IN THE CLAIMS

1-3.   (Canceled).

4.   (New) A system for delivery of high-performance online multimedia services comprising:

    a high-speed backbone, coupled to one or more network access points of a network, the high-speed backbone configured to receive general content and multicast content, the multicast content assigned to be multicast to a destination address;

    at least two regional servers, coupled to the high-speed backbone and including a first regional server and a second regional server which each provide a second level of caching for the general content, the first regional server configured to customize the multicast content to form a first version of multicast content and the second regional server configured to customize the multicast content to form a second version of multicast content; and

    a plurality of caching servers, including a first caching server coupled to the first regional server and a second caching server coupled to the second regional server, each caching server configured to provide a first level of caching of general content for a group of end user systems in a region served by the regional server to which it is coupled, and to provide the customized multicast content formed by the coupled regional server to an end-user system in the region.

5.   (New) The system of claim 4, wherein the high-speed backbone, at least two regional servers, and the plurality of caching servers comprise a private network.

6.   (New) The system of claim 5, further comprising a central server, coupled to the high-speed backbone, and configured to control the private network.

7.   (New) The system of claim 4, wherein the network to which the high-speed backbone is coupled comprises the Internet.

8.   (New) A method for delivery of high-performance online multimedia services comprising:

    providing a high-speed backbone, coupled to one or more network access points of a network, the high-speed backbone configured to receive general content and multicast content, the multicast content assigned to be multicast to a destination address;

    providing at least two regional servers, coupled to the high-speed backbone and including a first regional server and a second regional server which each provide a second level of caching for the general content, the first regional server configured to customize the multicast content to form a first version of multicast content and the second regional server configured to customize the multicast content to form a second version of multicast content; and

    providing a plurality of caching servers, including a first caching server coupled to the first regional server and a second caching server coupled to the second regional server, each caching server configured to provide a first level of caching of general content for a group of end user systems in a region served by the regional server to which it is coupled, and to provide the customized multicast content formed by the coupled regional server to an end-user system in the region.

9.   (New) The method of claim 8, wherein the high-speed backbone, at least two regional servers, and the plurality of caching servers comprise a private network.

10.   (New) The method of claim 9, further comprising a central server, coupled to the high-speed backbone, and configured to control the private network.

11.   (New) The method of claim 8, wherein the network to which the high-speed backbone is coupled comprises the Internet.